Title of Project: Cohorts of Women in Biographical Collections (CWBC)
Report Type: White Paper
Grant Number: HD-229059-15
Names of Project Directors:
     Alison Booth, Principal Investigator
     Worthy Martin, Co-PI
     Daniel Pitti, Co-PI
     Rennie Mapp, Project Manager
     Jeremy Boggs, Design Developer
Name of Grantee Institution: University of Virginia
Report Date: 12/01/2017. Submitted March 2018
Grant Type: NEH Office of Digital Humanities Level II Startup
Participating DH Projects:
     Collective Biographies of Women (CBW)
     Social Networks and Archival Context (SNAC)

Part I: Report

I.   **Narrative Summary**

Cohorts of Women in Biographical Collections (CWBC), a Level II Startup Grant project, has advanced the goals of the ongoing Collective Biographies of Women project (CBW). At the

same time, the Startup CWBC project has assisted Social Networks in Archival Context (SNAC) by laying the groundwork for its collaboration with "external resources," and by increasing SNAC's access to hundreds of printed biographies of women, including African American women. Further, the NEH-funded CWBC project has introduced more collaborative practices not only for CBW and SNAC but also for the digital humanities (DH) community at the University of Virginia.

CBW, directed by Alison Booth, indexes and analyzes networks of short life narratives from over 1200 print collections of chapter-length biographies of women of all periods and occupations, published in English, 1830-1940. Its distinctive approach to digitized narrative analysis, using a stand-aside XML schema, Biographical Elements and Structure Schema or BESS, will be briefly sketched below, as it will play a significant role in the next development of CBW, which will benefit from interconnection with SNAC's data. SNAC, directed by Daniel Pitti, aggregates descriptions of millions of lives from archives around the world, predominantly Euro-American males of more recent times. The work supported by this Startup grant, with two rounds of consultation with the advisory board, provides a model for research on prosopography, or networked, open-access, interoperable biographical profiles, in two quite different contexts: printed volumes and significant typological lists of women, largely in the public domain, and millions of archival descriptions associated with unprinted materials in thousands of repositories internationally. The CBW database interface is now public-facing and functional, while we approach the goal of clean data on more than 8000 persons and linked access, through WorldCat and HathiTrust, to the books or digitized texts. We have used two CBW cohorts to link hundreds of person records in CBW with a SNAC record for the same individual, establishing each set of resources as extensions from both projects.

This grant has allowed these projects to demonstrate the ways a researcher could draw upon both SNAC and CBW for expansive aggregated data on networks of historical personal records, or prosopography. SNAC, which was initiated even before there was widespread advocacy of linked open data, effectively mines and displays interconnections latent in archival descriptions. As libaries' and institutions' archives built digital portals to their holdings, they simultaneously created data about networks of persons and organizations of great interest to historical research. CBW explores the research potential of a rediscovered genre that itself built "documentary social networks" upon categories of difference, or what we term typologies; collections of women's lives presume the difference of gender but also sort tables of contents of short biographies by such typologies as nationality, vocation, and religion. SNAC searching seldom leads to narratives, and is preponderantly white male. CBW consists of English-language, all-female biography, with some selection of national and racial diversity but relatively little reference to the kinds of original documents linked through SNAC. As a result of the work funded by NEH for the CWBC collaboration, both research portals thus gain dimensions, and SNAC gains access points to women's lives that were narrated in CBW books and potentially interconnected with archival materials in other SNAC records.

The work in the grant period focused primarily on the CBW project and how to prepare its typological sets of person records for reconciliation with records in SNAC. The main steps of this start-up grant are listed here, with more details below:

A. Effective, substantial consultation with a distinguished international advisory board, first in a round of videoconferences, and second in usability testing, interview and questionnaire, and a symposium. Collaborations.

B. Redesign of the interface and enhancement of features and search functions for users of the CBW project site, as recommended by the board. Building community (social media, events).

C. Refinement of sets of cohorts of CBW persons: data cleansing on persons; typologies

D. Implementation of links between SNAC and CBW person records, and creation of new records in SNAC if CBW finds no "match" already existing in SNAC

E. Linking book records in CBW to library holdings and digitized texts so that users of both SNAC and CBW can pursue the printed narratives about the historical women

F. New models of teamwork with groups of interns and paid RAs, working on cohort data cleansing (III) and links to books (V) as well as BESS analysis

G. Outcomes: Presentations at the DH Conference, MLA Convention, invited lectures; blog posts; short or long articles

II.   Project Participants and Advisory Board

Principal Investigator Alison Booth, Professor of English, University of Virginia; now also Academic Director, Scholars' Lab, UVA Library (UVA)

Co-PI Worthy Martin, Associate Professor of Computer Science, Director of IATH (UVA)

Co-PI Daniel Pitti, Associate Director of IATH, PI for SNAC; now full-time Director of SNAC in UVA Library (UVA)

Project Manager Rennie Mapp, Lecturer in English; now Project Manager, Strategic DH Initiatives (UVA)
Design Developer Jeremy Boggs, Head of Research and Development, Scholars' Lab (UVA)
Vendors: Designer Anne Chesnut; Usability and Documentation Consultant Veronica Kuhn

Advisory Board:
Jean Bauer: Associate Director, Digital Humanities Center, Princeton University Library

Jason Boyd: Assistant Professor, Director, Texting Wilde Project; Associate Director, Centre for Digital Humanities, Ryerson University

Susan Brown: Professor of English, University of Guelph, Visiting Professor of English and Film Studies, University of Alberta; Director, Orlando

Jay Clayton: Professor of English, Director of the Curb Center, Vanderbilt University

Julia Flanders: Professor of the Practice of English; Director, Digital Scholarship Group; Director, Women Writers Project; Northeastern University

Regenia Gagnier: Professor of English, University of Exeter, UK

Natalie Houston: Associate Professor of English, University of Massachusetts at Lowell

Alan Liu: Professor of English, University of California, Santa Barbara

Laura Mandell: Director, IDHMC, Professor of English, Texas A & M University

Aimée Morrison: Associate Professor, Department of English Language and Literature, University of Waterloo

Bethany Nowviskie: Director, Digital Research & Scholarship, University of Virginia Library University of Virginia

Ruth Page: Reader in English Language, University of Leicester, UK

Sue Perdue: Director, Documents Compass, Virginia Foundation for the Humanities

Harold Short: Emeritus, Department of Digital Humanities, King's College, London

Gina Luria Walker: Associate Professor of Women's Studies, The New School for Social Research; Director, Project Continua

Additional consultation: Rafael Alvarado, Research Faculty, Data Science Institute

III. Work Completed

A. Effective, substantial consultation with a distinguished international advisory board:

1) **Fall 2015: Lengthy content and usability surveys of the advisory board**.
In order to collect usability data and responses to the intellectual content of CBW and its CWBC collaboration with SNAC, Mapp organized and conducted an extensive survey of the Advisory Board with Booth's assistance. Afterward, AdBoard members participated in four group videoconferences, aided by short guide in advance and Booth's slide presentation. These interactions formed the basis of Mapp's **CWBC Executive Summary** (See Appendix A).

The four two-hour sessions of small groups were lively discussions with such leaders in the field as Alan Liu and Bethany Nowviskie, as well as experts in our vicinity such as Suzanne Keen, Susan Perdue, Raf Alvarado and co-PIs who participated on site. The sessions and survey responses produced more animated enthusiasm, pointed insight, and helpful advice than we had even hoped. The board generously contributed their expertise and creative problem-solving abilities to all aspects of CWBC and CBW. Their most significant recommendations, drawn from Mapp's Executive Summary, can be described as follows:

a) Improve usability for search, tools, and navigation
b) Add readable biography texts to site

c) Focus work on cohort analysis prototype (CAP) to produce exportable and deployable cohorts from user search
d) Continue efforts to interlink CBW and SNAC

2) **Fall 2017-Winter 2018: Usability testing and conference**.
The second advisory consultation took the form of usability testing designed and monitored/recorded by DH usability consultant Veronica Kuhn, who recorded and reported on the AdBoard's interactions with CBW (see Appendix B); personal interviews; a questionnaire; a one-day conference with three advisory board members and the PI, followed by a half-day usability and collaboration planning session, for future work on the women writers cohort.

Feminist DH@UVA Symposium: Julia Flanders, Laura Mandell, and Susan Brown participated in this symposium organized by Booth and sponsored by John Unsworth and the Scholars' Lab. In a day-long session on January 11, 2018, attended by nearly forty faculty, students, and library staff, including collaborators on CBW and SNAC, each director of a feminist research project and center spoke about the technical and research designs of their latest work.

On January 12, Martin, Mapp, Kuhn, the distinguished guests, and Booth worked closely on 1) usability of CBW and 2) planning a collaboration, "Recovery 3.0" (tentative title), to coordinate all the bibliographical projects on women writers that are now facing end-of-career need of sustainable support.

This round of consultation with advisors was extremely productive, especially for future steps of the CBW project. It should be noted that the Usability Test and the January 12 session focused on BESS and the BESS Viewer as well as the interface, which have been ongoing developments of the CBW project. The achievements of the CWBC Startup phase in establishing links between CBW and SNAC and between CBW and WorldCat and HathiTrust were less the focus of this feedback.

CBW intends to complete reconciliation of all persons in CBW with SNAC. Based on the advisory board input, the next phase will be to fulfill the BESS experiment and explore its scalability and its comparative affordances with other approaches to computational narrative interpretation, primarily within the orbit of biographies of women.

B.   **Redesign of the interface and enhancement of features and search functions for users of the project site**

Prompted by fruitful written surveys and videoconferences with the advisory board, CBW improved its interface. Developers in IATH have contributed time, supported at a fraction of effort by this grant, to improving features of the site, search functions, and database infrastructure. In this period, Mapp and Booth worked closely with Jeremy Boggs of Scholars' Lab, experimenting with methods of rich-prospect browsing to make the project's collection of narratives and networks of persons more accessible and readily parsed. Mapp played a

leadership role and contributed weekly hours on this grant, but since early in 2017 her time was mostly dedicated to her new responsibilities in developing DH@UVA as a community, a website, and a graduate certificate program. This is one commendable outcome of NEH funding for digital scholarship: alt-ac careers for humanities PhDs. In May-July 2017, we rehired the designer Anne Chesnut to prototype the final design of the interface to adapt to the improved searching and visualizations designed by Boggs, Booth, and Mapp. We also hired independent contractor Veronica Kuhn (who also works for SHANTI at the University of Virginia) to consult on usability testing for the final consultations with the advisory board. Booth, Mapp, and Boggs collaborate, in meetings and on Github and Slack, to reconceive and rebuild the CBW site, making the forum and various visualization tools accessible, as advised by the board. This facilitates traffic between SNAC and CBW. The case study of Frances Trollope, designed by Mapp based on research by Booth, as well as the accessible categories of person typologies, along with other pathways into the site and the research project, are now largely implemented.

Community building and events:

The redesigned interface relates closely to the efforts to build a community and communicate research results. During the grant period, we designed and initiated a blog, entitled the CBW Forum, and have established Facebook and Twitter accounts. Research assistants and students have opportunities to contribute. Collaborations with organizations working in digital humanities at the University of Virginia include collaborative communications, and opportunities to co-author research results.

The March 2015 Symposium, entitled "Moving People, Linking Lives," was a success. Also supported by a grant from the Page-Barbour Foundation, "Moving People, Linking Lives" brought together researchers from UVA and elsewhere, some of whom are members of the advisory board, to discuss the challenges of research on prosopography or biographical cohorts in various disciplines and using diverse methods in the humanities, including DH.

In October, 2016, Booth, Martin, and Mapp (who was the primary organizer) held a one-and-a-half-day conference with three external speakers, Lauren Klein (Georgia Tech), Tanya Clement (University of Texas at Austin), and Chris Johanson (UCLA) , and rounds of lightning talks and discussions to help organize and conceptualize DH@UVA.

The Scholars' Lab hosted Michelle Moravec, history professor at Rosemont College, who has designed visualizations of CBW's data on historical women, for a workshop in 2016. Booth submitted an ACLS Digital Extension Grant application in January 2017, which would have entailed collaboration with Moravec and her students as well as a team at Washington & Lee University (finalist).

In January, 2018, the Feminist DH@UVA Symposium was also effective in building community as well as collaboration with the board.

C.   Refinement of sets of cohorts of CBW persons: data cleansing on persons; typologies

We refined our process of data cleansing and analysis of cohorts, working with teams of undergraduate as well as graduate researchers. For experimental "matching" with SNAC, we designated these cohorts (we note where LAMI student projects have been focused on particular cohorts). With varied teams of research assistants, undergraduate as well as graduate, we worked on cleaning the data on the persons in these cohorts (as well as other goals of the project): collaborators include Caleb Agnew, Lindsay Turner, Vanessa Braganza, Meghan Grumbling, Eva Lucy Alvarado, DeVan Ard, Neal Curtis, Jill Martiniuk, Julianne McCobin, Indu Ohri, Britta Rowe, Beth Sutherland, Valerie Voight, Christian Howard, Alison Thompson, Lloyd Sy, Wei Liu, Alex Lenkei, Jess Swoboda, in addition to students in several courses (Booth) and the summer LAMI program. http://cbw.iath.virginia.edu/people.php

- **Lola Montez siblings**. All members of tables of contents of books that include a biography of Lola Montez: their person records have been vetted; the cohort has been queried in the SNAC data for matching name/date and other attributes; the six most probable matches have been curated.
- **Female African Americans**. This cohort was assembled with the help of undergraduate researchers in the LAMI program, Kerwin Holmes, Jr., Malikia Johnson, Kayla Pinson, Joel Rhone, Lilybeth Shields, and Anthony Velazquez, and a UVA student, Eva Lucy Alvarado.
- **Presenters,** including biographers, translators, editors, and illustrators (many of whom are male); a team of the undergraduate and graduate students as noted above.

In progress: data cleansing, exploration of networks in the following cohorts:
All person records, including one-offs, will have at least one assigned type by April, 2018 (Booth and research assistants)

- Frances Trollope and women travel writers—Booth has studied the versions of Trollope using BESS, event analysis; Mapp and Boggs have designed a case study of Trollope
- Medical practitioners, partially vetted
- Latinas—LAMI students Gabriela Corona and Celeste Navas created some tables of contents and researched some individuals
- Queer subjects: detecting the unusually submerged cohort of non-heteronormative gender and sexuality—LAMI students Corona and Navas and students in Booth course researched this question, somewhat hidden in these collections
- One-offs, or the 6700 persons in CBW who have a single biography in the database, in progress; this list diminishes with elimination of duplicate records for some persons.
- Writers. We will be expanding the existing cohort of 1030 writers of all genres, as we finish typological assignments for the approximately 6700 one-offs. Many women famous or documented for another reason also published writing.

All cohorts can intersect in ways that support the argument of intersectionality (the co-dependence of differences of race, gender, class, and other identities), as when we discover an African American nurse, hypothetically one who has been honored as a lesbian in a website on lesbians of history, but who appears only once in a book in CBW's bibliography, and never in SNAC. As we complete the relational data of person records, chapters, and collections (adding missing tables of contents or even new collections to the post-1940 bibliography), new documentary networks emerge. We will turn to collaborations with other projects on historical women (not just women writers).

D. **Implementation of links between SNAC and CBW person records, and creation of new records in SNAC if CBW finds no "match" already existing in SNAC**

All above cohorts will be reconciled with SNAC, following the path that we paved with the Lola and African American cohorts. Of course, the ultimate goal is to have accurate person records throughout CBW that can be reconciled with SNAC in a batch.

The steps, broadly, are:
a) Cohort name-date patterns are queried in the SNAC database
b) the six statistically most probable matches are curated, to reconcile SNAC and CBW person records
c) if there is no existing SNAC record, a new one is created
d) CBW becomes an External Resource in that existing or new SNAC record
e) the CBW person page presents a link to the SNAC page.

The cohort of Lola Montez siblings has produced exciting results. Many of her siblings were high ranking and famous, though perhaps scandalous. SNAC frequently has multiple entries for such women. A recent focus on the cohort of Irishwomen in CBW highlighted, for Booth on the occasion of giving a lecture at UC Dublin, the representation of the Irish-born "Spanish" dancer in relation to others in the database. Matching the Lola cohort in SNAC has introduced some quite obscure figures such as Mary Clements, Lady Leitrim, who is a one-off in CBW, and now a record in SNAC, with the CBW book, Frances Gerard's *Some Fair Hibernians*, as the source.

As of February, 2018, the persons in a second cohort, African Americans in CBW's publications, have been reconciled with SNAC person records. Those 430 African Americans in CBW's database of collective biographies (subjects and presenters in the collections in the bibliography)

will either be introduced as new records in SNAC, as will Monroe Majors in the screenshot below, or will be added as an external resource link to the robust record in SNAC, as in the green selection of Phillis Wheatley in the image below. Notice that three of the six Wheatley results are more than likely to be multiples. CBW's mission is not to reduce the noise in SNAC, but the reconciliation of all persons in CBW will provide notes and the right-hand checked boxes to flag opportunities to merge SNAC records.

From the Reconciliation of the African American Cohort

The process of selecting the accurate SNAC record cannot simply rely on the algorithmic probability, not only because of the multiplicity of records for one identity in a resource compiled from many repositories, but also because many records in SNAC have very limited data, such as a link to one collection of papers, without more information about the firstname lastname pattern found in those papers. The many one-off, obscure persons in CBW's African American cohort (in contrast with most of the cohort of cosmopolitans and celebrities in Lola Montez's orbit) rarely enter the social networks of the named persons in library repositories. A researcher needs to probe a possible-match SNAC record for clues that the woman of a certain name could have been a woman of color known at the time of the publication of the collection. A case in point is a person who occupies both our African American and our medical women cohorts:

HOW BOOKS RESHAPE LIVES

Women / Persons ~ P23817

Dr. Ida Gray

Alternate Names:
- Dr. Ida Gray
- Ida Gray Nelson
- Ida Gray Nelson Rollins
- Ida Rollins
- Dr. Ida G. Nelson Rollins

Gender: Female

Life Dates: 4 March 1867 - 3 May 1953

Types:

African American
American (US)
Dentist
Medicine (doctor)
Member of committee or club
Member of local society

Note: Cincinnati. First African American Woman dentist. Married Sanford Nelson. President of the Professional Women's Club of Chicago.

[Find] persons with [1] degrees of separation from Dr. Ida Gray

Biographies (3 Total)

- Dr. Ida Gray
  Author: Monroe Alphus Majors . Collection Title: Noted Negro Women: Their Triumphs and Activities. 1893
- Dr. Ida Gray
  Author: Sylvia G. L. Dannett . Collection Title: Profiles of Negro Womanhood. 1964
- Ida Gray, D.D.S.
  Author: Lawson Andrew Scruggs . Collection Title: Women of Distinction: Remarkable in Works and Invincible in Character. 1893

Dr. Gray is one of only two women of the type *dentist* so far identified in CBW. SNAC had no record of her existence, but CBW books recognized her in 1893 and again in 1964. One 5.27% result, Ida Gale Gray, could be rejected with confidence upon investigating the sparse SNAC record. Ida Gale Gray was a member of the Daughters of the American Revolution.

Our cohorts, medicine, one-offs, and African American, overlap. Ida Gray emerges alongside a range of African American one-offs in Lawson Scruggs, *Women of Distinction: Remarkable for Works and Invincible of Character* (Raleigh, NC, 1893): at least 97 individuals in his reference work. The sibling networks are segregated by the tables of contents, with however an exceptional inclusiveness. We extend the scope of SNAC as we expand what we know about the collective biographies we have documented.

While Martin, Mapp, and Booth experimented with generating graphs of the networks of typologies, as a visual Cohort Analysis Prototype tool, we followed the board's advice that the visualization tool is not itself a research priority given the availability of easy-to-use, open-access visualization software. Cohorts (lists which any researcher may produce by searching persons by type) can be downloaded as CSV files, and available software (Gephi, D3) can be adopted according to the user's research questions. The abstract promised "the tools and process" and we have delivered the process, while enhancing the entire CBW usability and influencing the development of SNAC's relationships with collaborating resources. Meanwhile, the CBW project does provide radial graph visualizations of the networks formed by the tables of contents of collections, that is, of a cohort such as "Lola Montez siblings." Another visualization graphs the relationships of two "cousins," who do not share the same collection but share some siblings in common. Future work on CBW will enhance access and visualization of the varied typologies, persons, narratives and their textual features, and publications.

Further discussion of these cohorts and our research discoveries appears in Part II.

E. Linking book records in CBW to library holdings and digitized texts so that users of both SNAC and CBW can pursue the printed narratives about the historical women

Responding, again, to the advisory board's recommendation that the texts in CBW be more accessible to users, collection records now display links to search WorldCat (OCLC number) and HathiTrust digitized text of the book. For the curated cohort matches between CBW and SNAC, a researcher will have selected the suitable editions, and the links are set to find them. This fulfills the aim of leading SNAC users to the distinctive materials indexed by CBW.

Interlinking SNAC and CBW enables in-depth research of cohorts of female historical figures over time with archival networks of kinships, correspondence, and other affiliations. SNAC has expanded its representation of women with the new records drawn from CBW, particularly of the more obscure names in the African American cohort. The existing records in SNAC are improved by leading users to the rich interpretative narratives about these women; correspondingly, a user of CBW can discover archival materials and social networks that the original biographers, writing for a general audience, do not cite or never discovered.

F. New models of teamwork with groups of interns and paid RAs, working on cohort data cleansing (III) and links to books (V) as well as BESS analysis. Pedagogy.

Each summer since 2015, Booth has participated in the Leadership Alliance Mellon Initiative (LAMI) program for students from historically Black colleges and universities, Puerto Rico, or Latinas from the US to prepare them for graduate school; students are in residence at UVA for about six weeks, and are affiliated with the Scholars' Lab, which Booth now co-directs. Booth has mentored individual students, along with other faculty, staff, and graduate fellows. Some of these LAMI students are interested in biographies of women of color. Also during this summer session, Booth has coordinated paid teams of research assistants (graduate and undergraduate) who are enrolled at UVA. These RAs are either working on CBW's database—improving the data and person records for the cohorts—or on narrative analysis using the BESS schema. LAMI students have all expenses and a stipend paid. RAs are paid from research funds that Booth earned for mentoring in the LAMI program as well as by research funds Booth receives from the Library and the wage positions provided by the English Department.

We have made progress in the workflow and communication, preparing for further implementation of the project. Activities are coordinated by email, weekly meetings, Github, and Slack, as well as a password-protected administrative interface that allows RAs to receive and annotate their assignments and to find guides on their tasks. The "Lists" page helps teams search for the status of a person record or for cohort lists (http://cbw.iath.virginia.edu/cbw_db/lists.php). We use the SVN tool subversion server for contributing BESS work to the repository using oXygen. Documentation of the work process, largely written by Mapp and Kuhn, is migrating to the Knowledge Base in SHANTI, another research group at UVA.

Cindy Girard, XML specialist in IATH, introduces RAs to working with XML in the oXygen software, and Booth teaches them to analyze the narratives using the controlled vocabulary for element types. In several undergraduate and graduate courses since 2015, Booth has led assignments working with the data and the BESS schema in CBW. Leading the newly approved graduate Certificate in DH, Booth will offer a course focused on the CBW project.

G. Outcomes: presentations at the DH Conference, MLA Convention, invited lectures; blog posts; short or long articles

LAMI students, Booth, Mapp, Boggs, Martin, Eva Lucy Alvarado have presented our findings. LAMI students present to the UVA Library and at conferences in Boston and Connecticut. The CBW Forum and Scholars' Lab (with UVA news outlets) have represented the project, with associated social media. RAs have posted on the Forum.

Presentations re CBW since 2015:

1. Booth, Keynote, "Biographical Networks, Gendered Types, and the Challenge of Mid-Range Reading," Digital Cultures, Big Data, and Society Conference, University College Dublin, 15 February 2018; Emilie Pine and Margaret Kelleher, organizers.
2. Booth, "Can the Typologies Speak?: Problems of Feminist Biographical Networks," Feminist DH@UVA Symposium, 10 January 2018. One of four hour-long presentations/discussions of feminist DH projects.
3. Booth presented on the CBW project on two panels at DH 2017 in Montreal. She has collaborated with other researchers on the concept of reading scale (mid-range): Grant Wythoff, Sarah Allison, and Daniel Shore, in a panel accepted for DH2018 in Mexico City.
4. Booth, "Character Studies," Keynote, 6th International Conference on Narratology, Shanghai International Studies University, 21 October 2017.
5. Booth Lecture, "A Rhetorical Theory of Biography: Communal Voices and Types," Shanghai University, 23 October 2017.
6. Booth, Invited plenary panelist, "Strategic Typologies and Particular Webs: Digital Studies of Women's Lives," CUNY Victorian Conference, May 5, 2017.
7. Mapp, invited presentation, "Collective Biographies of Women: A Prosopographical Database with Added Narratological Analysis" at Biographical Data 2017, held in Linsk, Austria, sponsored by the Austrian Academy of Sciences.
8. Booth, Work-in-Progress Workshop, "Webs, Types, and *Middlemarch: Digital Studies of Women's Lives*," 25-pp. draft of essay submitted in advance; led discussion with faculty and graduate students at University of Utah. March 3, 2017.
9. Booth, panel organizer: The CWBC project formed part of a roundtable-style panel

on DH methodologies and diversity organized by Booth at MLA Convention, January, 2017, in Philadelphia; panelists Lauren Klein, Annie Swafford, Mark Algee-Hewitt, and Andrew Stauffer attended. Booth had to cancel due to an injury.

10. Booth, short paper, "Trollopini, Trollopized," on the Trollope cohort in Florence: NAVSA conference, Villa Pietra, Florence, May 2017.
11. Booth, Boggs, and Purdom Lindblad, Roundtable, "Design and Social Justice," Society for Textual Scholarship Conference, University of Maryland, June 2, 2017.
12. Booth, Panelist, "Frances Trollope and Harriet Martineau: Mapping Society in America Then and Now," NAVSA Conference, November 4, 2016.
13. Booth, Panelist, "Reading at Mid-Range and/or Named Entity Recognition," Narrative Conference, University of Amsterdam, June 15, 2016.
14. Martin, Boggs, Booth, Lindblad, and Mapp, prepared data and presentation: Sharing Biographical Data workshop July 11, 2016, DH2016 conference, Krakow. Boggs presented.
15. Booth, "Frances Trollope and Harriet Martineau: Mapping Society in America Then and Now," paper presented at NAVSA Conference, Phoenix, November 4, 2016.
16. Booth, Panelist, "Circulation of Histories: From the Oral to the Digital," Revising Revisionism Symposium, Brooklyn Museum, Elizabeth A. Sackler Center for Feminist Art and Project Continua, May 17, 2015.
17. Booth, Workshop and Planning Committee, Digital Diversity 2015: 20th Anniversary of Orlando, University of Alberta, May 7-9 2015. "Inside Collective Biographies of Women: Mid-Range Reading" 1-day Workshop Presenter.

Publications Related to CBW since 2015:

1. Booth, "Mid-Range Reading: Not a Manifesto," *PMLA* 132.3 (2017): 620-7, Theories and Methodologies forum on Distant Reading, Franco Moretti, respondent.
2. Booth, "Strategic Typologies and Particular Webs: Digital Studies of Women's Lives": essay invited for *Victorian Literature and Culture* in cluster on the Woman Card conference, due June, 2018; this began as work in progress shared with English Department, University of Utah, February, 2017; revision was delivered as lecture at plenary conference entitled "The Woman Card: Feminism and Victorian Studies, Past, Present and Future." CUNY Graduate Center, May 5, 2017.
3. Booth, "Lola Contains Multitudes," 15-pp. essay on Walt Whitman's lost novel, *Jack Engle,* and the character of the Spanish dancer, Inez, as a version of Lola Montez, a focus of research in CBW. Shared online in HC Commons, Academia.edu, and UVA institutional repository as well as the CBW and Scholars' Lab blogs.
4. Booth, "Character Studies," 1000-word response to James Phelan's target essay,

> "Authors, Resources, Audiences," under contract and forthcoming in *Style*. This advocates for the CBW project's strategic typologies and the study of nonfiction narrative.

5. "Prosopography and Crowded Attention in Old and New Media," *On Life-Writing*, ed. Zachary Leader (Oxford: Oxford University Press, 2015), 72-98.
6. Booth, "Prosopography," *The Encyclopedia of Victorian Literature*, ed. Dino Felluga, Pamela Gilbert, and Linda Hughes (Wiley Blackwell, 2015).
7. Booth, "Screenshots in the Longue Durée: Feminist Narratology, Digital Humanities, and Collective Biographies of Women," *Narrative Theory Unbound: Queer and Feminist Approaches,* ed. Susan Lanser and Robyn Warhol. Narrative Series. Columbus, OH: Ohio State University Press, 2015. 169-93.
8. Booth, "Frances Trollope in a Victorian Network of Women's Biographies." *Virtual Victorians: Networks, Connections, Technologies*, ed. Veronica Alfano and Andrew Stauffer. Palgrave Macmillan, 2015. 83-106. Digital Annex http://www.virtualvictorians.org/

Part II: Perspectives on CBW and the Collaboration with SNAC

This white paper serves both as a report and as a reflection on the meaningful coordination of typological, networked biographical records, or prosopography. It shares our discoveries so far and points to the potential rewards of this inquiry for interpreting the quantifiable human record. We are faced on one hand with overwhelming data about people of the past, and, on the other hand, with erratic preservation of formerly-registered recognition as well as the biased erasure of the records of most human effects. A great deal has been written and preserved, and yet most men and almost all women vanish with little documented acknowledgment. Many might assume that the rise of the Internet democratizes recognition, and that few human productions will fail to be registered in some way. Yet history and recent digital research both teach us to be skeptical about this assumption. The print-era efforts to comprehensively represent people and their cultural contributions—in encyclopedias and in the collective biographies that treat small sets of thematically related lives in more detail—have been followed by the efforts of libraries and databases to coordinate the preservation and dissemination of such biographical information. Computers are aiding the coordination of equitable acknowledgment in forms that might last. But of course the problem of prosopography remains daunting.

Digital exchange and analysis of data on persons, events, and documents is still an intellectual and technical challenge. Researchers and archivists alike desire centralized access to resources upon which authoritative biography would be based; they seek ways to parse associations revealed by guided queries and to measure what's missing. Many lives are unrecorded due to illiteracy, exclusion from property transactions, or posthumous neglect, but women (customarily changing names upon marriage) are particularly vulnerable to these elisions, especially before the modern era or across linguistic differences. In 1938, Virginia Woolf lamented the absence of biographies of housemaids in Britain's great dictionary of biography.[1] CBW's bibliography

---

[1] "It is much to be regretted that no lives of maids, from which a more fully documented account could be constructed, are to be found in *the Dictionary of National Biography*." (*Three Guineas* [London: Hogarth, 1938], chap. 2 n. 36; Project Gutenberg of Australia, 2002, http://gutenberg.net.au/ebooks02/0200931.txt).

reveals the construction of cohorts of women's lives by publishers and biographers who aimed not to recover the lives of housemaids but to share representatives of different female types, far beyond the mothers, wives, mistresses, or correspondents of famous men that tend to surface in archived papers in libraries, and hence in SNAC. *Cohort*, in this context, refers to a set of named individuals connected by a common factor, which may be factual or interpretive. Cohorts that emerge from CBW's collections are categorizations of representations, whether within one book or in a sample corpus (all the books that include a life of a particular person), or alternatively in search results for different attributes such as occupation. CBW already has identified all 1271 collections by type or organizing principle, so each book can be viewed as a cohort or set. Another kind of cohort can be framed through a search by a person's name, a common initial query in many resources. Thus, a search for "Joan of Arc" in CBW leads to her person record listing 68 collective biographies published at a varying pace across the decades (1855 [9 books] and 1917 [8]). These results consist of a set (cohort) of names tied to each of the chapters in this corpus of 68 printed books, effectively, Joan's "siblings." In SNAC, Joan is linked to those who wrote about her such as Mark Twain, and users can trace an archival network.

Scholars recognize that even the most skeletal account of a person is highly interpretative, and that biography and history are forms of narrative. These insights motivate two ongoing studies of persons in networks to model programmatic, joint inquiry into repositories and textual interpretation. CWBC has compared means of access and interpretation of portrayals of women in different research platforms and online references (such as Wikipedia), yielding a more textured understanding of both the portrayals and the women portrayed.

CBW studies all English-language collective biographies of women, representing many different periods, nationalities, and occupations rather than women writers of a single tradition. Exceptional in its focus on the form of nonfiction prose, it develops digital methods for interpreting narrative structure in all the texts in that genre, using BESS. Such *mid-range reading* enhances the historical research and network analysis that we can pursue in the highly structured representation of types and groups in publications identified by date and place. Not least, CBW is an open-access platform for research on more than 8500 persons and 14,000 narratives collected in 1271 books published primarily 1830-1940, with improving tools for visualizing networks of persons and increasing links to the coordinated archival access provided by SNAC. Whereas encyclopedias and reference works outline historic facts, family or other relationships, these narratives offer versions intended to entertain and instruct general readers. From the page for each collection in CBW, a researcher can link to WorldCat (OCLC number) and HathiTrust (URL for digitized text) to read these fascinating narratives. Users of CBW's database now can readily identify cohorts or sets of named persons (by occupational, bibliographical, and other controlled vocabulary of types) and download CSV files for further visualization.

SNAC aggregates some 3.7 million descriptions linked to holdings in approximately 3,000 repositories. SNAC's Prototype History Research Tool displays information about people and affiliations extracted from descriptions of archival documents, most unpublished. Users of SNAC can locate records of individuals, organizations, and families; browse featured descriptions; and discover networks. Persons are associated with each other in SNAC due to their co-occurrence in descriptions created by archivists. We have found that the CBW cohorts have served as effective queries of the SNAC data, revealing the potential of this intersectional

research to share discoveries about the world's archives and a genre that encoded women's roles. Our comparative certainty about the materials referenced and accessed through our database can amplify the presence of marginalized people in a powerful research portal that draws upon many repositories with diverse metadata standards. Crosswalks to the holdings and records of institutional and family affiliations in SNAC ensure that CBW is not an enclosed greenhouse of rare specimens or perennial favorites.

Our methods and tools will be adaptable to further research collaborations and agreements of this kind. In the immediate outcome, users will be able to harvest information—life dates, nationalities or associations, languages used, occupations—about the subjects of chapters in CBW's books and uncover relationships beyond the horizons of written versions of those lives while taking advantage of SNAC's versatile links to such resources as Wikipedia, ArchivesGrid, and DPLA. At the same time, CBW has much to offer an authoritative cooperative of archives.

Screenshot detail of Lola Montez record in CBW, showing SNAC link, typologies, and alternate names, and the corresponding record in SNAC, showing the CBW link.

snac Q Search

Montez, Lola, 1818-1861 Alternative names

Dates:
Birth 1818-02-17
Death 1861-01-17

Nationality: IE
Language: French, Undetermined, English

Biographical notes:

Irish dancer, actress, lecturer, writer, and political activist, Lola Montez was also known as Marie Dolores Eliza (or Elisa) Rosanna Gilbert, Maria Dolores Parris Montez, and Comtesse de Landsfeld.

From the description of Lectures and correspondence of Lola Montez, 1842-1950 (inclusive), 1842-1858 (bulk). (Unknown). WorldCat record id: 612379070

Born Marie Eliza Gilbert in Ireland, Lola Montez was an internationally-known dancer who performed in San Francisco and various mining towns in Calif. from 1853-1856, with a brief tour in Australia in 1855-1856.

From the description of Lola Montez letter : Munich, to Monsieur le Baron : ALS, 1846 Oct. 22. (California Historical Society). WorldCat record id: 122385303

The British actress was famous for being the mistress of King Ludwig of Bavaria.

From the description of ALS, 1859 June 15 : 26 Park Lane, [London], to Mr. Dubois. (Copley Press, J S Copley Library). WorldCat record id: 16956870

Lola Montez (real name Elizabeth Rosanna Gilbert), Irish-born dancer and actress.

From the guide to the Lola Montez manuscript material : 1 item, ca. 1843?, (The New York Public Library. Carl H. Pforzheimer Collection of Shelley and His Circle.)

Dancer, b. Marie Eliza Gilbert.

Links to collections

Archival Collections — 43
Related Resources — 19
Related External Links — 5

Virtual International Authority File
Wikipedia
WorldCat Identities
LC Name Authority File
Collective Biographies of Women

Related names in SNAC

People — 27
Families — 0
Organizations — 3

Lola Montez has been a focus of a corpus of books that include a biography of this adventuress, and her cohort of "siblings," the women also featured in these books, is now established in SNAC. SNAC gathers a much longer list of alternative names for Lola Montez than CBW does, and while it turns up such material as an Australian scrapbook, it draws upon existing archival descriptions which may promulgate a contestable iteration of an identity; the Library of Congress and other SNAC sources give her birth date as 1818 instead of 1821. Under "Related Names in SNAC," 27 names are "associatedWith" Montez, not one of which is yet known in CBW's research. The list of 27 omits her famous lovers such as Franz Liszt or Ludwig I and all the proper names that recur in short and long biographies of Montez. Researchers following the Related External Link from SNAC to CBW would find a different approach to Montez among varied cohorts communicating how the person was (or should be) perceived. Anyone interested in biographical history can gain a more wide-ranging perspective on thousands of women's lives, ranging from well-documented to little known, ambiguous, or legendary. (Is the Little Captive Maid the same as the Little Hebrew Maid in collections of women of the Bible?) Our synthesis of assets and approaches to the study of people in sets or cohorts will be increasingly useful for prosopography and other digital projects.

The second cohort or experimental set of women, African Americans, have been more laborious to curate. Throughout the work on typological cohorts, but especially in this phase, we have been reflecting on the perils of typologies, arguing that they can be strategic and handled with care. To label the women of color in our database is undoubtedly to repeat profiling. Perhaps it is excusable because the women have been so categorized in the original club memberships or educational institutions such as historically Black nursing schools, and hence in the printed books that we have included in the CBW bibliography. The options of multiple typologies help to restore some complexity to the person's representation.

NOTED NEGRO WOMEN

THEIR TRIUMPHS AND ACTIVITIES.

BY

Mr A. MAJORS, M. D.

———

"A race, no less than a nation, is prosperous in proportion to the intelligence of its women."

———

The criterion for Negro civilization is the intelligence, purity and high motives of its women.

———

THE HIGHEST MARK OF OUR PROSPERITY, AND THE STRONGEST PROOFS OF NEGRO CAPACITY TO MASTER THE SCIENCES AND FINE ARTS, ARE EVINCED BY THE ADVANCED POSITIONS TO WHICH NEGRO WOMEN HAVE ATTAINED.

———

An African American, Dr. Monroe A. Majors, assembled *Noted Negro Women* and published it with Donohue & Hennberry, a large publishing house, in 1893 (access to publishing was an issue). It is not widely available in the original, but it was microfilmed, and a facsimile was printed from that. This facsimile is now digitally accessible on HathiTrust, and through a link from CBW. The book perfectly matches the "our women" rhetoric of so many of the CBW texts: "a race, no less than a nation, is prosperous in proportion to the intelligence of its women," as the epigraph on the title page affirms. Negro women have attained advanced positions, which proves "Negro capacity to master the sciences and fine arts." The incapacity of women to reach the highest levels in any field was widely declared at this time, but so was the shameless wondering aloud if African Americans could go further than trade school. The image above shows three of the four epigraphs. Monroe's list of 112 women includes Phillis Wheatley, Frances E.W. Harper, and other notables, but also many now obscure.

Amanda Berry Smith has been listed in both CBW and SNAC before our work began on matching the African American cohort, but in a sense she is a discovery for this team, a result of focus on this cohort. A SNAC search for Amanda Smith yields eight potential results, narrowed by the middle name and life dates (1837-1915) to two or possibly three archival identities that ought to be merged with the substantial record id 65034743. Smith had not become familiar to us after years of studying African American History Month prosopographies. Within the horizon of Majors's 1893 African-American list, Smith ranks with several of the most eminent women of color of all time, but she was an evangelical preacher—not a type that lasts into the present. Smith is in six CBW books, five of these all-African-American, one a volume of Christian exemplars of all eras. Below is the beginning of the chapter by Marshall W. Taylor in Majors's *Noted Negro Women*:

AMANDA SMITH.

Missionary, Author, Minister.

WHAT is the place of Amanda Smith in American history? Has she any place there? Mrs. Smith is an historic character. The biography of great women, and especially great women of the Negro race, would be sadly deficient without her.

Of this race in the United States, since 1620, there have appeared but *four* women whose career stands out so far, so high, and so clearly above all others of their sex that they can with strict propriety and upon well established grounds be denominated *great*. These are Phillis Wheatley, Sojourner Truth, Frances Ellen Watkins Harper and Amanda Smith.

More than a score of Negro women have arisen to the heigths of fame which leaped beyond the bounds of the States in which they have resided. Nor is there a single State in our country, North or South, but that could point to Negro women, a score or so in number, who are zealous of good works, endowed with a noble spirit and a love of race, sex and self which is truly praiseworthy and distinguishing. Indeed, in every city, village, or country neighborhood, a leading Negro woman, who is a full match for its best and leading Negro man, can be found.

Screenshot of Marshall W. Taylor, "Amanda Smith," in Majors, *Noted Negro Women* (1893), ch.98

As noted, this is much the same "our women" rhetoric, with all its racialist and nationalist trappings, as is used by contemporary white-dominated prosopographies. Like social scientists of the turn of the twentieth century, these presenters identify the progress of a few named representatives as an argument for the progress of a race. No one at the time questions consensus on an elite of greatness, or a reliable single narrative of history. Yet the third paragraph above opens a window for social history and female equality: "every city" has its "leading Negro woman."

Cohorts overlap, and collections may reveal selective principles that measure prejudices and social differences over time. For instance, CBW includes two collections of female nurses published in 1928 and 1929 in the US: *Makers of Nursing History* by Meta Rutter Pennock (ca. 1892 - 8 January 1959), who, according to our research, was "not a nurse, nor African-American. Barnard BA 1917; Columbia MA in public health education, 1941. She spoke at 1924 meeting of Ntl Assoc. of Colored Graduate Nurses. Edited *The Trained Nurse and Hospital Review*." This note attempts to account for the biographer's appearance in a chapter, "History of the National Association of Colored Graduate Nurses," in Ada B. Thoms, *Pathfinders: a History of the Progress of Colored Graduate Nurses*. We are currently examining the 109 biographies of the nurses in Pennock's "unmarked" collection—no doubt Florence Nightingale and Clara Barton, named within the first ten biographies, ring a bell (the contents appear to have some chronological but no alphabetical order). The fourteenth chapter is about Adah B. Thoms (alternate spelling of name), author of the segregated volume, *Pathfinders*. How many colored nurses appeared in Pennock's extensive list? We have some investigation ahead of us, beginning of course with reading the volume, which is still in copyright. I had assumed that the grand title, *Makers of ...History* might filter out women of color. But clearly Pennock endeavored to promote the progress of nursing education for women of color. Perhaps the most segregated lists in CBW are those dedicated to the advancement of colored people (coeval with the NAACP).

Appendices

Appendix A

I.   CWBC Executive Summary

Distilled Expertise and Creative Solutions from Our Advisory Board

By Rennie Mapp, Project Manager
February 8, 2016
Sources:
Summarized from advisory board's written surveys and videoconference transcripts, November 2015, and the successful September 2015 Proposal for an NEH Level II Startup

Introduction
CWBC Advisory Board members bring impressive expertise to their analysis of CBW and its collaboration with SNAC. In November, board members willingly completed complex written surveys and hours of videoconferencing, demonstrating their commitment to the project as well as the relevance of CBW's biographical and narrative resources to the spectrum of scholarly

fields represented on the board, including Digital Humanities, biography and life-writing studies, gender studies, and narrative theory and narratology (among others).

In particular, Board Members devoted remarkable creative energy to identifying strengths and solving dilemmas within the project. Their energy enables this report to add valuable strategic direction to the remaining term of CWBC's NEH grant. In written and oral comments members offered enough detailed recommendations and specific encouragement to produce not only a **roadmap** but also unambiguous **user profiles**, **use cases**, and **site architecture** to guide our development efforts at IATH. CWBC, CBW, and SNAC are very fortunate to have this team of advisors.

Main Advisory Board Recommendations
1. Improve usability for search, tools, and navigation
2. Add readable biography texts to site
3. Focus work on cohort analysis prototype (CAP) to producing exportable and deployable cohorts from user search, rather than on building social network visualizations
4. Interlink CBW and SNAC

Overview of Meeting and Survey Results
Board members frequently admired the "richness" of CBW's data resources, and appreciated the progress made in approaches to visualizing and searching within those resources. They identified the quality of CBW's metadata as a core strength. They genuinely look forward to using CBW's tools for narrative analysis and social network analysis in their research and teaching.

Imperfect usability was the most frequent criticism. As users, board members seemed to feel that access to CBW's riches is just out of reach due to its disorienting and incomplete interface. They offered precise, creative solutions with the obvious hope and expectation that the project can remedy its UX problems in the near future, so that researchers and students will be able to access CBW materials with ease. These **usability problems** fell into three categories:
- search functionality
- tool manipulation
- navigation

Although the board sees CBW's content as especially strong, there were two main recommendations for added **content**.
- Most importantly, board members want to see the texts on which the site's data and metadata are based. Because they currently cannot read the women's biographies that underlie and inspire the project's insights, they feel that one of the most potentially rewarding aspects of the project is missing.
- Secondly, they look forward to the augmentation of both CBW and SNAC that interlinked person records will bring.

CWBC's advisory board has offered clear recommendations not only for the CBW-SNAC collaboration but also for the CBW project site on which the collaboration largely depends. They endorse CBW resources as valuable for both research and teaching, and seem eager for the site's

improvement. Many if not all of their recommendations can be implemented before the end of the NEH Digital Humanities Level II Startup Grant period.

Strategic Priorities Going Forward

1. For the NEH as funding agency, CWBC's value lies in that it proposes to--
   - 0collaborate at IATH to match and interlink person records from CBW and SNAC, and
   - develop related social-network cohort analysis tools.

   Quantitative and qualitative analysis of the results of the CWBC Advisory Board survey and videoconferences reveal that **improving CBW's UX is crucial to both of these NEH-sanctioned goals.** CWBC advisory board members are intellectually motivated and fairly tech-savvy; yet they struggle to perform basic functions within our searches and tools, and feel frustrated by missing navigational context. As members made clear, in collaboration with SNAC, CBW has the potential content to be a widely-used and recognized scholarly and pedagogical resource, but UX difficulties stand in the way of that potential. **UX improvement must therefore be as high a priority as the development of new tools and interlinked databases**.

2. **Cohort analysis tools should depend on user-collected searches based on typological metadata** as well as **on documentary social network analysis**, as board members make clear. This is an important refinement of the goals of CBW's social network analysis instruments. There are several options that may help connect users' searches in CBW with visualization tools:
   - Before the visualization tools are complete it may be necessary to develop a conduit or pipeline tool/methodology to mediate between CBW search results and CWBC cohort visualization tool(s).
   - Alternately, a dashboard tool for users (off-the-shelf or developed in-house) could mediate between search results and cohort analysis tools.
   - A savable user sandbox could serve as the source from which network tools could draw cohorts for analysis.
   - Improved search functionality and UX would be an indispensible foundation for this conduit tool.

3. A prototype process for **matching person reco**rds between SNAC and CBW, with an accompanying index of the likelihood of exact-match accuracy, has worked well (see Robbie Hott's recent matching work).
   The CWBC team needs to determine next steps--
   - to match all of CBW's vetted person records with SNAC, and
   - to define and map out the process for **interlinking these records** between CBW and SNAC.
   - Again, integration of interlinked person records between SNAC and CBW is a UX issue as well as a technical challenge.

What does the Advisory Board Value Now?
   - **Project approaches**

- o Of CBW's five types of approach to its biographical and narrative content, members clearly rated the **BESS Viewer** and the **social network analysis tools** highest
- o Curated cohorts, database search, and bibliography also received positive mentions
- **Data usefulness**
  - o Members ranked **BESS markup**, Women/Person **Search**, and CBW **typologies** fairly equally in usefulness
- **Metadata usefulness:**
  - o **Typologies** were clearly ranked highest, followed by **BESS markup**, **social networks**, and **person records**
- **Search experiences**
  - o User search narratives based on survey assignments were mostly ranked as **unsuccessful** experiences overall, although users cited revealing results and potential value as points of success along the way
  - o Problems frequently cited were
    - Lack of consistent, contextual **navigational aids**
      - Most users expressed a desire for these aids to be universal, meaning that they would like every page to indicate the user's position in the site as a whole
    - Lack of **pre-population** in search fields
    - Lack of **data context**; ie., lack of an understanding of the "contours" of the amount and number data included in the project
      - Prose explanations of data context were widely rejected as a way to offer these contours, in favor of navigational or visualization features
    - **Tools** such as the social network r-graphs and BESS Viewer were **hard to find**
  - o Respondents sometimes cited their awareness of the danger of overcrowding search and tool pages, but expressed a preference for more contextual information rather than a "clean" page design

What Added Features Does the Advisory Board Suggest?

- o Access to **readable text versions** of the chapter-length biographies was the most frequently requested feature
- o More extensive **faceted searching** was the second most frequently requested feature
- o Replacement of unique identifiers on search-result pages with more revealing **link titles** was also frequently mentioned, along with strong recommendations for **search-result pages to contain more data and metadata**
- o The project **home page** was politely but widely seen as problematic. Suggestions included--
  - Use of **images** to represent project approaches (rather than text)
  - Foregrounding of a link to the project **Forum**, with automated updating of latest posts

- - - Clearer, less opaque establishment of **the "contours" of CBW's data and metadata**
    - **Clearer statement of what CBW is** (see appendix section on NewCity website consultation for more advice on this)
    - Substitution of the word **"Persons" for the "Women"** search category
    - Help in understanding the search **category of "Biographies**" (also see NewCity appendix)
- **SNAC collaboration**
  - Members greeted the news of the success of the first matching of batch person records appreciatively, especially because the matching process promises to improve the representation of underrepresented groups within SNAC (and potentially in Wikipedia)
    - They questioned how the process of matching records would lead specifically to **greater representation of CBW persons within SNAC**
    - They expect **CBW person records to appear within SNAC**
    - They strongly suggest CBW person records to contain a link to matched SNAC person records whenever possible, and vice versa **(interlinking)**
  - Members also cited SNAC's navigational context as friendlier and more usable than CBW's
- **BESS Viewer**
  - Although most respondents reported **confusion about how exactly to operate** the BESS Viewer, they recognized it as a highly valuable tool
  - They especially recognized the value of the **handmade, stand-aside markup**, which they tended to mentally associate with the more general value of CBW's database typologies—all seen together as **beneficial typological work**
  - Board members suggested a **navigational aid** that would enable users to search for texts with associated BESS markup **according to the typological metadata for persons, chapters, and collections**
- **Social network cohort tools**
  - Board members applauded the analytical and historical value of Booth's concept of visualizing Documentary Social Networks based on co-occurrence of persons within and across related collections.
  - Most did not find the RGraph tool to be user-friendly or revealing due to low functionality
  - Function requests:
    - More **readable person and collection titles**
    - The option to **eliminate collections as mediating nodes** between persons
    - A **matrix based on collection or person type** in order to visualize the variety among women in a single collection
  - Some members expressed lack of familiarity with the concept of the cohort, and suggested **examples of pre-packaged cohorts** as a way to educate users and make tools easier to understand

Roadmap for Improvement:

1. Quick wins for **User Interface** based on Advisory Board suggestions
   a. **Project Home Page**: more visual information; less prose clutter

      b. **Faceted search pages** that "reveal the contours of the data"
      c. A **navigational guide** (widget?) that always locates and orients the user within the site architecture.
      d. More transparent, linked, navigable **search results**
2. Quick wins for tools
      a. **BESS Viewer** is almost great
          i. Provide more instruction (possibly tooltips) for BESS Viewer
          ii. Make buttons at top less inscrutable
      b. **Documentary Social Network R-graph**
          i. Make it possible to eliminate collection nodes
          ii. Research other R-graph tools (D3)
          iii. Replace collection numbers with titles
      c. **Cohort Analysis Tool**
          i. Develop means to capture and save search results
          ii. Develop tool to visualize them as networks
3. Quick wins for **CBW-SNAC collaboration**
      a. Match all vetted CBW person records
      b. Add links from matched SNAC person records to CBW (under SNAC's "Links to collections" section)
      c. Add links from matched CBW person records to SNAC person records

Project Development Materials, Based on 2015 Survey and Videoconferences
1. **Site architecture diagram**, as foundation for navigational aid
2. Four **user profiles** (aka personas) based on board members
3. **Use-case descriptions** (diagrams and prose descriptions) to aid programmers in search and tool development
4. **Sequence diagrams** to aid programmers in search interface development
5. Agile user stories to articulate specific **feature requests** and changes within larger development categories

Anticipated Problems with this Roadmap
1. Lack of UX expertise
2. Lack of front-end developer time
3. SNAC team focus on database transition
4. Lack of tool developer time

Appendix B

CBW Usability Test Report

January 2018
By Veronica Kuhn
- Summary

Five members of the Collective Biographies of Women Advisory Board completed a set of tasks on http://cbw.iath.virginia.edu/. I recorded their interactions and impressions, which are summarized in this report. I also calculated completion rates and error rates for each task.

The test indicated:
- Users can easily use the website to find, explore, and manipulate the content of person records, chapter-biography records, and book records.
- The database is large enough so that users can find information on the specific women of their choice.
- Users had questions about the methodology behind CBW and what they were seeing.
- Despite users' interest in the data, the interfaces of the BESS and social network tools irritated and confused testers.
- Users were intrigued by BESS, and they wanted more ways to manipulate the data, in particular by correlating and tracking controlled values.

From these results, I recommend:
- Rework the model for methodology content
- Use basic usability heuristics to improve BESS and social network tools
- Help users create visualizations from BESS data that correlate controlled values and life stages

I've also included a full list of small-scale suggestions from testers, and from test observations, at the end of this report.

Methods

I tested five members of the Collective Biographies of Women Advisory Board using the remote conferencing software Zoom. Members were asked to complete tasks while speaking out loud about their thoughts or concerns. You can see the full list of tasks and moderator script on [this document](#). I recorded their interactions and analyzed the video afterwards. One tester didn't attempt a few tasks because of time constraints. In this case, I jumped to the BESS visualization, since that area was high-priority for analysis.

There were no technical issues with setting up the test or recording using Zoom, and the method was free. I recommend this method for future usability studies.

All testers were told the data was completely anonymous, that I was an independent contractor responsible for collecting honest feedback, and that I would aggregate comments in the final report. As such, testers are labeled P1-5 through this report, and I've included no full transcripts. I have, however, included quotes or paraphrasing to illustrate key observations. I've also included a [full list of suggestions](#) on the last page, which includes every suggestion testers made.

For each task, I computed two metrics:
1. *Completion rate* represents the percentage of testers that completed the task successfully. If they had to go on to the next question, the task was marked as incomplete.
2. *Error rate* represents the percentage of testers that committed errors while completing a task. For example, if they went to **Biographies** instead of **Persons** first, but then realized they needed to go to **Persons** and successfully completed a task, that would be an error.

- Results

Exploring the homepage

Three testers used a "trial and error" method, where they decided to try out the tools immediately to learn more about Collective Biographies of Women. They clicked on most of the icons on the homepage, and tried a few test searches. Two of these testers mentioned they were trying to gain an idea of the "scope" of the project.

Two testers went to the "About" link on the top menu bar to read more about the project.

Exploring the BESS tool
All testers struggled with showing any markup on the interface at all. Three testers could not
     show all of the BESS markup without guidance. Testers that successfully showed markup
     did not use the toolbar: instead, they clicked on the area next to the text. Two testers
     looked at the text navigator in the top right corner of the page for guidance: to them, this
     was the clearest "visualization" the tool offered.
P1 and P3 wanted a more "systematic" way to view the data, and in particular to correlate
     controlled values with life stages. P3 suggested the text overview in red at the top right
     corner include different colors for different controlled values.
Task completion
"Completion rate" represents the percentage of testers that completed the task successfully. If
     they had to go on to the next question, the task was marked as incomplete.

	Completion rate (%)	Error rate (%)	Notes
Find a person	100	0	
Find a list of related women	80	20	
Visualize a social network relationship	**50**	33	All users ended up on the persons page, but some people didn't see or understand the link to the tool. P2 and P3 looked to the homepage to find out how they could visualize relationships using CBW, but couldn't find the explanation or link to the tool. P4 thinks it would be hard for students to understand what was being represented.
Filter network visualization	**75**	33	P1 took some time to figure out they needed to click "Display R-graph." P2 and P4 mentioned that "Person: 1", "Collection: 1" isn't easy to understand, and that grammatically these weren't distances.
Find a list of women in the same biography	100	0	
Find a list of persons of a type	100	0	
Find a multi-subject	**33.3**	0	I suspect this was a failure of wording on the

person-bio			part of the test. Users also struggled with realizing that one biography could have multiple subjects.
Use person dates	100	0	
Find a biography of a type	100	**75**	P1, P3 went through "Persons" first, before realizing the difference between a "bio-type" and "person-type." P4 didn't notice that biographies had a type filter at first.
Find publication information	100	0	
Find a link to the full text	**50**	0	P2, P3 didn't think to use the search filter. It's as if they just assumed that most text wasn't available through the chapter records, since the previous records they had viewed didn't include full text.
Find BESS	**40**	0	Once testers found a chapter record that contained BESS, completing this task was easy. However, some didn't consider using search to find which chapters included BESS.

A good benchmark "completion rate" for usability tasks is above 78%. This is the average completion rate for tasks, taken from a study of 1000 usability tasks using a wide range of software. In red, I've highlighted the completion rates below these numbers. Of course, certain types of tools may have higher or lower standards, but this is a good rule of thumb for us right now.

Positive feedback

More tasks were successfully completed than unsuccessfully completed. Database retrieval tasks (finding people, filtering people by types, and generally manipulating records) were simple for most tasks and most users.

Additionally, every woman users searched is represented in the database in some shape or form. The bibliographic metadata itself brings novel insights -- for instance, P1 mentioned it was interesting that Ouida was the only novelist included in one collective biography, over other novelists that might be more familiar to contemporary readers.

Barriers to task completion

While I've made a few notes on hangups for each task with low completion rates or high error rates in the table above, there were several recurring barriers across test participants and tasks.

Bug: The interface isn't displaying my search results.

When users enter text into a field, they are offered autocomplete options. If they do not click on the autocomplete options, but instead enter the text manually, searching displays "No

results."

Because this was not a usability error, but more of an interface bug, it's not included in the statistics above. Nevertheless, this bug should be fixed as soon as possible. Furthermore, the search should return results for "fuzzy" searches. For instance, searching "women" instead of a full title in the Books title field currently returns no results.

I don't understand what I'm seeing./I'm unclear about the methods.
P1. "There's nothing here that explains how degrees of separation work."
P3. "It's not clear how people are related... what does one degree of separation mean?"
P4. "I'm looking for a list of all entities, which might help me..."

Users had questions about the way certain elements were generated, in particular the typologies and the list of women. Many times, they either did not reference or did not find the existing documentation on the CBW website.

This tool's interface makes me feel dumb.
P4. After seeing the BESS visualization for the first time and not seeing any markup: "I think I've done something wrong. Let me go back [to the biography page]."
P2. After playing around with the BESS visualization: "It makes me feel stupid."
P3. After trying to figure out how the date filter works on the network visualization: "I'm actually feeling pretty stupid right now."

Users expressed significant frustration with the BESS visualization and social network tool. For both tools, they were frustrated that the data wasn't immediately apparent. They were also frustrated by the labeling of options, which they thought were opaque. Finally, interacting with the visualization options was perplexing. For example, P2 and P1 were confused by the date options in the social network tool (why is "before 1941" an option?).

Recommendations

In light of the results of testing, I suggest the following:

- Help users understand methods by **rethinking the model for methodology content**. This includes centralizing the documentation and providing documentation within the individual tools themselves.
- Try some **simple usability heuristics** to improve the BESS and social network tools. Any redesign should:
  - Make labels jargon-free
  - Use interactions familiar to users
  - Clearly indicate "empty states" and visibly respond to user input
- Help users **create more visualizations** with BESS data

On the following pages, I provide a detailed explanation of each recommendation, with some potential examples.

Rethink model for methodology content

The information the testers wanted on methods already exists on the CBW website on the various approaches pages, case studies, and the **About** page. However, users didn't reference this information for two reasons:

- They didn't take the time to search for it on the website (P1, P5), or
- It wasn't in the location they expected (P2, P3).

CBW's database model is clearly-defined: the approaches literature could use similar definition. I recommend a two-pronged approach of *centralizing* this information, while also *providing methodology in context.*

Centralize methodology pages

Right now, most of the methodology is available on the homepage through the various approaches sections. However, testing showed that users:

- Expect a consistent content model. The "Narratology" page includes a description of the BESS schema, but no link to the tool. The "Typology" page includes links to typologies, but no description of the rationale behind them.
- Expect centralized and independent location. They want a one-click solution for finding all documentation from any page.

While the content model can be reworked, I suggest:

- Include all documentation through the "About" page.
- Make sure every approach includes both a description and a link to the feature.

Provide methodology in context

Describing methodology in its context means providing the answers users need at the moment they're asking the question, instead of making them browse the **About** page or other approaches sections.

This doesn't necessarily mean a lot of extra content, or even instructional text. Content can be pulled off existing pages. This also isn't meant to be a bandaid for usability issues. We have to recognize that CBW is a rich tool, with rich approaches, and these approaches should unfold as users explore the tools themselves.

The search pages are a good example of methodology in context: for instance, BESS is clearly defined.

For an example of how this could be put to use on other pages, take the **Typologies** page. Right now, it's a list of person types. Some testers weren't immediately sure what these typologies were. Other testers immediately argued that these typologies were restrictive and problematic. Of course, CBW currently provides answers to these questions and rebuttals to these concerns on the **Case Studies** page. In the example below, I've simply excerpted a paragraph from the **Case Studies** page above the list of person types:

Typologies

For each person record, we provide a list of person types, or "P-types," which we derive from careful, "slow" reading of the person's chapter-bios. These types do not impose our

---

> own 21C categorizations of women, but instead attempt to capture the categories in which biographers and editors have positioned women in the past.
>
> **Adventuresses, Heterosexuality, Beauties**
> - [Adventuress, self-promoting entrepreneur](#)
> - [Beauty, celebrated](#)
> - [celibate, chaste or virginal](#)
> - [Coquette](#)
> - [Lover or mistress of great man, middle or lower rank](#)
> - [Prostitute](#)
> - [Sexual agency, transgression (hetero)](#)
>
> **Relation To Male: Lover, Muse, Mistress**

This puts the appropriate methodology where users are likely to need it the most. It also provides a strong argument in support for CBW's methods, by addressing reader's potential questions from the get-go.

You can imagine similar text within the BESS and Social Network tools, on chapter pages, and more. I'd suggest:
- Determine where "snippets" of methodology content could fit into the existing CBW framework.
- Develop firm style standards for this text to keep it as succinct as possible.

Try usability heuristics for tools

Improving the BESS and social network tools can seem like a monumental task, especially given users' frustration. However, simple usability concepts, drawn from existing methods in usability design, may be able to relieve frustration.

Show interface feedback

When a user opens a tool and sees no data, it can be disconcerting. They aren't sure whether there's a bug in the interface, or whether they've done something wrong. Whenever possible, make sure a tool opens on *some* data, like all of the markup, or a social network graph with a person-distance of 1.

Avoid novel controls or buttons

If possible, it's best to use tried-and-tested methods of showing and filtering information. This helps users understand the interface by referencing past experience on other websites.

Many users could not show all markup using the existing BESS tools. Here's the current interface:

CBW - TEI/BESS Analysis

Collection: a020 - Calamity Jane and the Lady Wildcats
Biography: a020#bio03 - Belle Starr

Show: ⦿ All ⚪ None ⚪ Text ⚪ Selected

Stages of Life: ▮☐ Before ▮☐ Beginning ▮☐ Middle ▮☐ Culmination ▮☐ End ▮☐ After

While labeling plays a role in the confusion here (what is none? what is selected?), there are a few additional issues:

1. Radio buttons are traditionally used to select *mutually exclusive* values. However, these current labels aren't mutually exclusive options. For example, a user might expect that showing "All" would also show "Text" -- but then why is there an independent "Text" option?

2. *Selection* only happens once users can see what they're selecting. It's unintuitive to make that a first-level option, especially one that's mutually exclusive, when users need to see the markup (by clicking "All") before they select.

On most websites, checkboxes are usually used instead of radio buttons for selections where multiple options are possible. If there are many options, some sites may offer a "Select all" or "Clear all" button. It's reasonable to assume that checkmarks in this situation would be more intuitive to users:

Collective Biographies of Women

Collection: a020 - Calamity Jane and the Lady Wildcats
Biography: a020#bio04 - The Extinction of a Comet: Lola Montez

Show: ☐ Chapter text ☐ Markup
Stages of Life: ▮☐ Beginning ▮☐ Middle ▮☐ Culmination ▮☐ After

(show) Word Cloud   Selected Text

Along the same lines, work with user's natural instinct to display markup, *then* select specific controlled values. This example on the next page would show up only when "Markup" is selected.

Collective Biographies of Women

Collection: a020 - Calamity Jane and the Lady Wildcats

Biography: a020#bio04 - The Extinction of a Comet: Lola Montez

Show: ☑ Chapter text ☑ Markup

Markup: [Show selected markup only] [Clear selection] [Select all]

Keep styling consistent between interactions

In CBW, there are three different types of buttons to access features:

Type 1	Explore Radial Graph Display Publication Bar Graph
Type 2	[**View BESS analysis**] [**Find**] persons with [1 ⬍] degrees of separation from Belle Starr
Type 3	[Text] [Text + BESS]

Keeping all these buttons consistent will help users "learn" the site, which in turn helps them find the features they need. I suggest keeping all buttons in the "Type 2" style, since this type is easiest to find and click. "Type 1" was the worst type: at least two testers missed the buttons entirely.

Don't use jargon for labels

Labels should always use terms that target users understand immediately. Some labels in CBW are very easy to understand: "Annotated Bibliography" and "Publication Graph" are 'exactly what they say on the tin.'

However, in other instances, label jargon prevents users from discovering a key feature. This was a particularly significant barrier when testers were trying to find the social network, a task which only had a 50% completion rate, even when testers visited the person page. A "radial graph" might be more readily understood as a "publication network" or "social network."

Current Name	Suggested
Show R-graph	Show Network Refresh Network

Explore Radial Graph	Publication Network Social Network

Help users manipulate data

BESS was interesting to users, and is a really powerful tool. However, several users wanted to expand the capabilities. In particular, they wanted to correlate and visualize the trajectory of terms. Here are some use cases they suggested:

- *I want to see whether "x" controlled value appears most often at "y" life stage.*
- *I want to be able to compare whether controlled values "a" and "b" usually appear together.*

For these users, an "ideal" BESS visualization, as I understood it, would immediately support a narratology argument with a single screencap. With so many controlled values, this is certainly a non-trivial data visualization and usability challenge. However, it also provides a rich opportunity for CBW to put innovative BESS data at scholar's fingertips.

2018 Usability Test Appendix

List of people that testers searched

Dolly Madison
Mary Hayes
Hypatia
Ouida (x2)
Margaret Cavendish
Aphra Behn

Full List of Suggestions

	Suggestions from testers	Suggestions based off interactions
Person record	• Sort the chapters by year • Include sources for alternate names	• Make the tool links more visible • Highlight the information to the left of the chapters in some way (by using a slightly darker colored background, for example.) • Link typologies to the list of women with that typology
List of relationships	• Provide options for filtering and sorting the list	

	• Provide methodology behind the list	
Book record	• Include links to all person-records included in the volume • Show publication information at the top, instead of the bottom	• BESS + Text should link to the new BESS visualization • Make sure "Text" and "BESS" buttons match the buttons for other tools (see full ux report)
Biography record		If there's no text, alert the viewer to the fact, then provide a link back to the search results with "Full Text" selected. This helps viewers understand that there IS text out there, even if that volume doesn't have any text!
Person search	• Provide search by start and end dates, instead of one "Year Lived" field • Include more in-depth typology filtering	• Make the "name" field respond to search that isn't the full name. • Make sure users can get search results even if they don't click on the suggested autocomplete
Book search		• Make the "name" field respond to search that isn't the full name. Almost all users searched a fuzzy term like "women"
Biography search	• Let users sort results by date	• Make the "name" field respond to search that isn't the full name. Almost all users searched a fuzzy term like "women"
Typology page	• Include methodology on page • Typologies are potentially problematic	
Cohorts page	-	• Add a drop shadow or some other way to indicate the images of the women are clickable
About page	• Include methods of selection	• Provide a link to the Case Studies page

	• Include scope • Include information available on "Case Studies" • Include a "full list of entities" • Make sure the "CBW is…" list has a parallel structure • Make sure "stand-aside schema" is a known term	
BESS Methodology Page	• "Narratology" does not prepare them for the content of this page • Expect some way of visualizing the text or narrative	• Include a direct link to all the BESS texts available (this will likely be results from "Biographies" with the BESS option selected)
BESS Tool	• Provide more ways of viewing the data (ie, "X term appears most commonly at the Y stage of life." or "x and y often appear together.) • Clarify and simplify controls at the top • Make it obvious how the Event markup elements fit together	• [See report](#)

Network visualization tool	• Make controls more intuitive. In particular:
	   ○ "R-graph" makes no sense
	   ○ The "Distance" fields make no sense
		• [See report](#)

Other	• It would be nice to have a centralized search function, so that they can just search the entire database at once	

Wireframe of BESS Usability Recommendations (Kuhn)

I've summarized some of the suggestions below into the wireframe above. This includes:

- A way to search for controlled vocabularies in the BESS analysis
- A way to view a "higher level" visualization of selected vocabularies
- A "term first" approach, where users search and select the terms that they're interested in. This might be a different view than one that lets you read the text with every BESS controlled vocabulary beside it, which might look like AB's original design.

L.Mandell:

- On homepage: didn't immediately see you could scroll down to see all options.
- Wanted a way to see the all the BESS controlled vocabulary terms included in the text.
    - Could we also include a BESS vocabularies
- Wanted a way to search for paragraphs with the controlled vocabularies within a text
    - Points out that this search would have to be "fuzzy" to some extent, just to account for the fact that people might search "legal trial" instead of "trial in law court", etc.
- Thought the text selection options at the top level were opaque
- Thought the ability to see the "tree" of terms was important
- Some references for inspiration: PRISM, The Poetess Archives (for search sidebar)

J.Flanders:

- New ways to visualize text encoding are tricky…
- Suggests having instructional text as popovers within the interface
- To solve the "multiple color" problem, suggests letting people choose tags first, then view the selected tag as a symbol next to the text. This limits the amount of information on the page, which makes it easier to display.
    - Suggests there is pedagogical value in having students experiment this way. For example, they might guess "Hmm, looks like this paragraph contains a rhetorical question. Let me see if I can select 'rhetorical question' from the list to see if it displays on the paragraph." It becomes more of a back-and-forth between the interface and the person.

Additional consultation with advisory board at conclusion of Startup:

In November, 2017, Booth prepared a summary of the SNAC-CBW reconciliation process and linking, as described in this White Paper, and shared it with members of the board who had not been able to participate in the usability testing. This summary was sent with the following preamble of questions to these members. We received one response (Regenia Gagnier) and held two extensive in person interviews, with Bethany Nowviskie and Alan Liu; the other responses were relayed during the Feminist DH@UVA Symposium (Mandell, Flanders, Brown).

Update on NEH Level II Startup: Cohorts of Women in Biographical Collections

Note: this is prepared for consultation with advisory board members in anticipation of the white paper.

Please read this synopsis. At the end, you will see the outline of the usability test and questionnaire, added for your information. We have a set of test results from some members of the board. I

am not asking you to take that test or answer the questionnaire. It's up to you if you want to interact with these.

We intend to apply for further grant support.

1) Which of the initiatives and topics below would you recommend for the **priorities of the next phase** of the project? Assume that some of the implementation can occur with current (limited) funds for small teams of research assistants and the limited contributed time of Boggs, Martin, Mapp, and Booth. If this outline of the project's status does not include the goal(s) you would recommend, please state your recommended priority/ies here.
2) Explain your recommendation in 1.
3) CBW plans to continue to collaborate with SNAC, Wikipedia, and to explore collaborations with the projects led by Susan Brown, Julia Flanders, and Laura Mandell. Please discuss potential collaborations, or point out projects that CBW's participants should view as models for next phases.
4) CBW contributes to narrative theory of nonfiction and digital textual studies. We anticipate that topic modeling and other methods of textual analysis would be fruitful for CBW's corpus, as a point of comparison with BESS analysis. Would you endorse the idea of testing distant and mid-range reading in this way?
5) Please freely raise questions or share comments you have about the project, its usability, and its aims or achievements thus far. We will try to schedule a conversation to follow up on your response.

Gagnier's response, written, is summarized here:

Priorities should be the model of LAMI and RA Pedagogy and the work with the Typologies of women in the database. "As I haven't been able to use the resource successfully for my own research (see 4 below), I nonetheless wholeheartedly support the training of graduate students in sophisticated digital methods. I am also interested in the knowledge that comes out of interrogating taxa and typologies."

On potential collaborations: re pedagogy: "Dino Felluga's COVE project has employed (through our universities) graduate student copy editors who are simultaneously trained in DH. Our students at Exeter love it, and from the Uni's perspective it is cost effective." On question 4, Would you endorse the idea of testing distant and mid-range reading in this way? "Yes, of course, per your PMLA article."

For research on persons, this usability experience was negative: "In trying to use the CBW site, I have tried unsuccessfully to trace a network of women poets from the British 1890s, using names like Alice Meynell and "Michael Field." I want to know how interconnected they were and how they theorized poeisis, whether there was anything like a poetic community. Can CBW offer me anything? If so, how?"

Both Liu and Nowviskie had read the memo, and in separate interviews of over an hour they affirmed the importance of the BESS experiment with narrative "reading." Liu raised two main headings of concern: **"scalability"**—the team work in selecting typologies (by which he apparently meant the BESS vocabulary) requires labor, and cannot be extended easily to the entire corpus or more extensive investigation of biographies or other narratives. This needs to be evaluated and considered. For CBW to fulfill its aims, however, it **must be validated by such scaling up.** That is, our subset of narratives and networks of persons seems to confirm its own data unless it is aligned with norms; trends in gender representation within this database need correlation with other data. Liu also was recommending the correlation of BESS methodology with other forms of textual analysis.

Nowviskie also concurred with Booth's plan to compare findings of the BESS experiment with textual analysis such as topic modeling. **You have a training set** was the similar locution both Liu and Nowviskie used. While CBW does not plan machine learning experiments in quite the usual sense, it is useful to consider 250-300 completed BESS analyses as a basis for testing algorithmic approaches to discovering textual features in the corpus.

Appendix C: From the Grant Proposal and Appendices: Illustrative Analysis of CBW's and SNAC's Rich Data and Explanation of BESS and BESS Viewer

CBW's Experiment on Cohorts of Women in Biographical Collections

A prime insight of CBW is that biographies follow conventions and cluster in groups or types. (BESS and the CBW database identify persons by a controlled vocabulary of types, usually occupational; the same person may be represented in multiple types.) CBW enables a reception history of types of women and the ways that groups of women were represented relative to each other. The genre of book that we study makes the social tropes of biography explicit and ripe for interpretation. The books' contents may be a "world history sample" (*World-Noted Women, or Types*

of Womanly Attributes of All Lands and Ages) or may select person types (*Noted Negro Women*, *Mothers of Great Men*). Users interested in religion may find the Quaker preacher Elizabeth Fry in 54 books, including *Quaker Women* (1915), or discover 12 books with titles similar to *Heroines of Methodism* (1857); 34 collections are typed "sectarian (religious group)."

Consider one Mrs. John Livingstone. She appears in only three CBW collections, forming a cohort (a set of associated names in a sample corpus) with 37 other women.[2] To visualize this network, Fig. 1 is centered on Mrs. Livingstone with the three collections positioned on the innermost circle and the other women of those collections around the second circle. Of her 37 collection "siblings," nine also appear in these same three collections and nowhere else in CBW, confirming some consensus on the names that represent "Notable Women of the Scottish Reformation," a specific historical episode in one location. Other members of Livingstone's documentary "siblings" range into other collections and in Figure 2 those additional collections are positioned on the third circle. In these more eclectic tables of contents, some of the Scottish heroines of religious conflict intersect with a multi-national set of women widely recognized today, including writers such as Harriet Beecher Stowe and heroines of war such as Joan of Arc or the Countess of Montfort. Mrs. John Livingstone has 331 documentary "cousins."[3]

Scholars in the humanities may wish to pull out other threads in female biographies besides religion. Thus our disparate cohorts for CWBC include occupation, nationality, and periods. The women who are associated with Italy (136; not all were born in modern Italy) form an interesting period or temporal cohort as well, since Italy became a unified state in the later nineteenth century, centuries after early modern women rose to cultural prominence in Venice, Florence, and elsewhere. These women raise occupational issues as well: many were once notorious and are now admired as pioneering women of letters. For example, Tullia d'Aragona, "a 16th-century Italian poet, author and philosopher" according to Wikipedia, barely surfaces in SNAC through her relationship to Tasso and Girolamo Muzio's publication of her dialogue on love. In CBW, however, Tullia's limited reception can be studied in a graph of Publications Per Year on her Display Page: only one Victorian version (published in 1859, during the Italian Risorgimento), followed by a cluster in 1909-1929, a high point of demand for collections of the types "adventuresses," "beauties," "mistresses, courtesans." For an example of this period of books on "adventuresses," the collection titled *Enchanters of Men* (first published in 1909) includes Tullia with Lola Montez, the opera singer Jenny Lind, and Napoleon's younger sister Pauline Borghese in a largely Franco-Italian list of 23. World-traveled Lola Montez, a node in one of our first sample corpora, also belongs in Calamity Jane and the Lady Wildcats (1927), a parade of notorious women of the American West, alongside suffragette Carrie Nation and convicted horse thief Belle Starr, among others.

---

[2] Janet Fleming Livingstone or Mrs. John Livingstone surfaces in a cohort of late-seventeenth-century Scottish Dissenters; many are wives of Presbyterian ministers or widows of martyrs or exiles. This relatively persistent cohort includes noblewomen, but most are historically obscure (life dates or first names unknown, calling for matching with SNAC identities and other records). Two collections have versions of the same list of persons: a029, Rev. James Anderson, *The Ladies of the Covenant: Memoirs of Distinguished Scottish Female Characters, Embracing the Period of the Covenant and the Persecution* (London: Blackie, 1850), with reprints through 1880; a157, William Chapman, *Notable Women of the Covenant: Their Lives and Times* (London: Swan Sonnenschein, 1883). The third book including Mrs. Livingstone is a068, Rev. Donald Beaton, *Scottish Heroines of the Faith: Being Brief Sketches of Noble Women of the Reformation and Covenant Times* (London and Glasgow: Catt; Adshead, 1909).

[3] One near relation (two degrees of separation) of Livingstone is Sister Dora, who illustrates the exponential possibilities: she appears in 20 collections with 141 "siblings" and then a prodigious 3645 "cousins."

IV.   As noted in the narrative, sample corpora consist of texts of all the CBW books that include one nodal persona; we have focused the BESS experiment on six sample corpora centered on Sister Dora, Lola Montez, Caroline Herschel, Cleopatra, Frances Trollope, and Charlotte Corday. We can type the personae, the collections, and the networks of relation among them. Sister Dora (see Fig. 5) appears in twenty collections ranging from queens and writers to missionaries, from 1880 to 1930; many of the 141 women (including repeats) took charge of the souls, bodies, appetites, or discipline of the poor, often dressed in uniforms *à la* Salvation Army. Heroines of this sort become obsolete with women's access to the professions; Sister Dora drops out of sight. Lola Montez (see Fig. 3) appears in fourteen collections also published in decades surrounding 1900, with 151 other women, some several times. Cleopatra, a node in her own corpus of 32 books, shares a book with Montez three times. The Egyptian pharaoh appears in combinations of history and legend, power and feminine error (from a 1937 scholarly study on ancient queens to collections of queens including the living Victoria). A search in SNAC for Cleopatra yields 717 results, most of which derive from the Shakespearean femme fatale. SNAC's user can also link to libraries, archives, and online entries on Cleopatra, but since she long predates most archival collections and modern historiography, it is not surprising to find "biographical notes not available for [her] record." The SNAC record would invite more informed research, possibly discover more archival documentation, and generate wider interest if it also led researchers to CBW's sample corpus for Cleopatra.

Caroline Herschel joins in Sister Dora's cohort (in 4 books together as siblings), but she is in "2 degrees of separation" from Lola Montez (see Fig. 4; eleven women "connect" as cousins with both Herschel and Montez). These disparate women who never met or corresponded can be studied in cohorts sorted by perennial roles and networks formed by of interrelated documents, persons, and events: learned women, mothers and sisters of famous men, aristocrats and queens, writers, Frenchwomen, heroines of heroic violence including suicide. Herschel, a much-awarded astronomer, Bavarian spinster, sister of a famous man, and one-time vocalist in Bath, is in 25 books. Trollope, the novelist, travel writer, failed entrepreneur, and unhappily married mother, is in 9 collections, while Corday, the French aristocrat who assassinated Marat, is in 22 books. BESS analysis of these sample corpora will be complemented by CWBC's linked data of SNAC archival descriptions and the improved interactive visualizations in Cohort Analysis Prototype.

FIGURES:

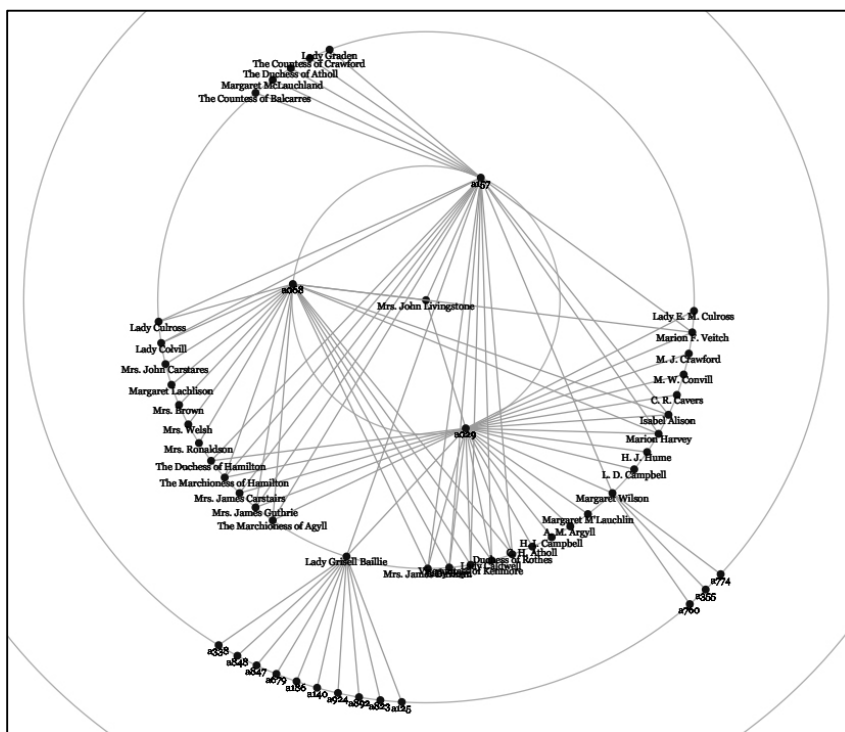Figure 1. Mrs. John Livingstone and 37 documentary siblings.

Figure 2. The 16 collections containing Mrs. John Livingstone and her documentary siblings.

Fig. 3 The Same Person Differs from Herself. Lola Montez in different artists' conceptions: in a557 Mayne, *Enchanters of Men* (1909)*;* a788 Terhune, *Superwomen* (1916). Though often photographed, Lola has a chameleon instability in the many portraits that imagine this notorious celebrity.
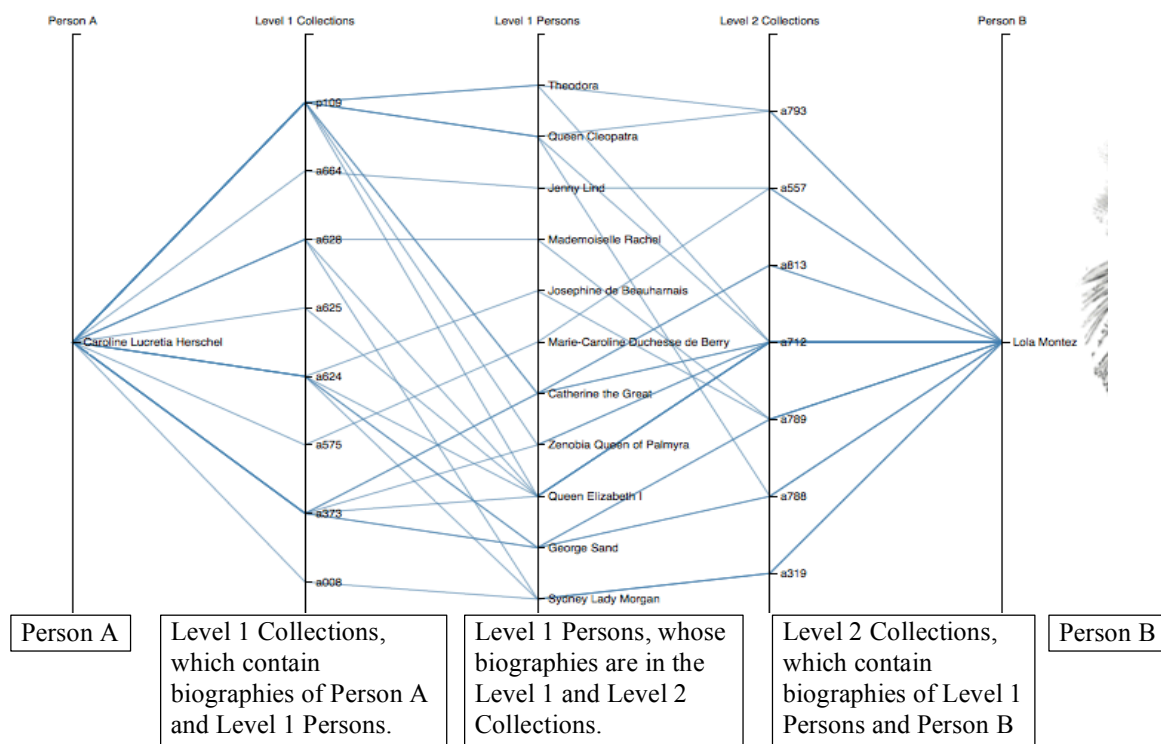
A: Caroline Lucretia Herschel ~ B: Lola Montez

Fig. 4 This shows two degrees between Caroline Herschel (A) and Lola Montez (B). Herschel shares eight collections with eleven historical women who are also listed in seven collections with Montez. Cleopatra shares one collection with Herschel but three with Montez.

Fig. 5 Different People Take on the Likeness of Their Cohort(s). Portraits of Sister Dora (nurse, hospital administrator) and Caroline Herschel (astronomer) in a034 Green-Armytage, *Maids of Honour* (1906). The book's twelve subjects show that unmarried women have succeeded in a range of careers (as the subtitle *philanthropy, nursing, poetry, travel, science, prose* suggests). Very different individuals can resemble each other in the same medium, style of image, and context. While SNAC's Prototype presents portraits captured from

Wikipedia, CBW has a "library" of image files of portraits from public-domain books and a prototype of a carousel of versions of the same person.

SNAC Sample Visualization

Sample force-directed graph of Susan B. Anthony SNAC network (courtesy of Brian Tingle, California Digital Library)

Anthony appears in 19 collections in CBW, including a001, Abbot, *Notable Women in History* (1913), which she shares with nodal women Cleopatra, Herschel, and Trollope and 69 other women (the sample corpora centered on these nodal personae will include this biography of Anthony). Needless to say, the American Anti-Slavery Society, in the above graph, is not a person in CBW, but Abraham Lincoln is (as a joint subject of a chapter in *Our Presidents and Their Mothers*). But the contemporary suffragists and reformers clustered in the upper right of this image also cluster in the table of contents of *Notable Women in History* and other books: Elizabeth Cady Stanton (in 13 CBW books, 8 of them with Anthony); Lucy Stone (in 7 books, 5 of them with Anthony); and Lucretia Mott (16, 11 with Anthony). A search for the Blackwell Family (top blue dot near label, above) leads to Lucy Stone and a description that notes that she "kept her own name after her marriage to Henry Browne Blackwell in 1855." All these American women activists are associated in the collection of the Alma Lutz papers that includes a calling card of German-British astronomer Herschel, who died in 1848 unaware of the Seneca Falls Convention, the foundational event of women's rights in the U.S.

BESS and Next Steps

Mid-range Reading and BESS. CBW has evolved a unique approach to textual interpretation at mid-range, between the big data mined from all texts in a genre and the close readings of a few

classic novels, stories, or films that have dominated narrative theory.[4] The PI, co-PIs, and PD, with consultant Suzanne Keen and research assistants, devised an innovative XML stand-aside schema, Biographical Elements and Structure Schema (BESS). BESS editors identify element types occurring in each paragraph of a short biography: *Stage of Life*; *Events*; *Persona Description*; *Discourse*; and *Topos*. The growing set of BESS files allows analysis of the conventions and variations in structure, narrative technique, and biographical detail within corpora (versions of one woman and her cohorts) and across types of women in cohorts and networks, as well as types of collections. Some 230 detailed analyses of narrative chapters in highly indexed publications offer a kind of training set for other forms of textual analysis. BESS, simplified or adapted as needed, can become the model for comparable projects and serves well for teaching close narrative analysis as well as the concepts of prosopography.

Lola Montez, read across the grain of the versions of her life that she initiated, generates significant reflections on the type of nationality, for instance in the following paragraphs, shown in the BESS Viewer with selected editorial choices for paragraphs 2-4 of Frances Gerard's *Some Fair Hibernians*, a patriotic collection that takes pride in women who were celebrated in the English metropolis and elsewhere.